34 Topological analysis of metabolic networks. Application to the metabolism of *Mycoplasma pneumoniae*

T. Pfeiffer[1], T. Dandekar[3,4], F. Moldenhauer[3] and S. Schuster[2,3]

[1]Friedrich Miescher Institute, Theoretical and Computational Biology, P.O.Box 2543, CH-4002, Basel Switzerland

[2]Humboldt University Berlin, Institute of Biology, Invalidenstr. 42, D-10115 Berlin, Germany

[3]Max Delbrück Centre for Molecular Medicine, Dept. of Bioinformatics, D-13092 Berlin-Buch, Germany

[4]Biocomputing & Structures Program, EMBL, D-69012 Heidelberg, Germany

Introduction

The topological analysis of biochemical networks has recently attracted increasing interest in the light of functional genomics [1–4]. Methods for the computer-aided synthesis of biochemical pathways have been developed earlier (for review see [5]). One of these approaches is based on the concept of elementary flux modes [6,7]. This approach has been presented at two earlier BTK meetings [8,9]. It allows one to test whether sets of enzymes form a coherent pathway allowing mass balancing for each intermediate and complying with the directionality of reactions given by thermodynamic constraints. Importantly, pathway analysis can be performed without the knowledge of kinetic parameters.

Here we will demonstrate how elementary mode analysis can be used as a tool in the study of bacterial metabolisms. On a small scale, guidelines in reconstruction of metabolisms based on pathway analysis could already been given for several micro-organisms such as *Haemophilus influenzae*, *Mycoplasma hominis* and *Methanococcus jannaschii* [4,7]. As a specific example, we will here consider *Mycoplasma pneumoniae*, the genome of which has been completely sequenced [10]. *M. pneumoniae* is a parasitic bacterium living in the human respiratory tract and a causative agent of tracheobronchitis and primary atypical pneumonia (cf. [11]). With a genome size of only 800 kbp, it is one of the smallest self replicating cells.

Methods

Elementary modes can be readily computed by our computer program METATOOL written in C [12]. The program, which is available from ftp://bmshuxley.brookes.ac.uk/pub/mca/software/ibmpc, uses a list of reaction equations and a declaration of reversible and irreversible reactions and of internal and external metabolites as the only input. Metabolites are classified as internal if they have to fulfil the steady-state condition. External metabolites, in contrast, are imported from, or excreted to, the surroundings of the cell, accumulate in the cell (e.g., storage metabolites) or are present in large excess (e.g., water).

For analyzing large metabolic networks it is necessary to have special tools to handle the problems caused by the high complexity of such networks. For programs analyzing the structural properties of networks, such as GEPASI [13] or METATOOL [12], the running time and memory usage increase rapidly with increasing system size. Other problems caused by a high network complexity are limitations in visualization, tractability and comprehensibility. To cope with these problems, we choose the following strategy. First, the metabolic network is constructed from internet sources. For a better analysis, it is then decomposed into smaller subnetworks. After analyzing their topological properties, the interplay between them can be studied.

The subsystems have to be delimited from other parts of metabolism by considering additional metabolites as external. This concerns central branching points in metabolism, that is, metabolites involved in a high number of reactions. We found that it is a useful rule of thumb to consider as external, in addition to source and sink metabolites, all those substances that participate in more than five reactions. These can be detected automatically by a C program written by one of the authors (T.P.) based on an algorithm from graph theory for analysing the connectivity of networks. However, the interpretability of results is improved if this classification is slightly edited afterwards. Another reason for this editing is that the data downloaded from databases are sometimes inconsistent or not perfectly up-to-date (cf. [4]).

The database KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.ad.jp/kegg/kegg2.html, see also [3]) is a useful source of data about metabolic networks in numerous organisms. It offers not only a visualization of metabolic systems, but also various useful files, which can be downloaded by anonymous ftp.

In the ENZYME database being part of KEGG, information about enzymes is available. The database follows the EC nomenclature and contains EC number, enzyme name, and the catalyzed reaction. Since not all side reactions are mentioned and the names of the reactants are often inconsistent or have a different degree of specificity, parsing that database is almost impossible. The COMPOUND database contains information about metabolites. It is useful to find aliases for metabolite names. The REACTION section contains data about the (enzyme-catalyzed) reac-

tions. On the ftp server, also species-specific data are available, mainly of species with completely sequenced genomes. By parsing the enzymes with EC numbers in the COMPOUND and REACTION databases, it is possible to construct an input file for METATOOL.

The metabolism of *M. pneumoniae* involves enzymes of glycolysis (leading to the products acetate and lactate) and sugar interconversion (including parts of the non-oxidative pentose phosphate pathway), nucleotide interconversion, fragments of lipid metabolism and a part of arginine degradation. This parasite is not able to synthesize amino acids or nucleotides *de novo* [10]. Thus, it is a model organism for the determination of minimal genetic requirements of an autonomously reproducing cell.

For constructing the metabolic network of *M. pneumoniae*, data from the database KEGG have been downloaded. According to the procedures described above, a list of the enzymes (EC numbers) was used to generate a list of reactions [14]. Furthermore, the enzyme lists obtained were also cross-checked by data from the literature (cf. [10]) and sequence analysis methods [1,4].

Results and discussion

Using the procedures described in the previous section, the metabolism of *M. pneumoniae* can be decomposed into 40 subnetworks. 21 of these can be deleted because they only contain external metabolites, that is, they involve single disconnected reactions, which are usually irrelevant side reactions of unspecific enzymes. The remaining subnetworks correspond to, among others, sugar import, glycolysis and pentose phosphate pathway, lower part of glycolysis, nucleotide interconversion, inosine-hypoxanthine interconversion, one-carbon unit pool and arginine degradation [14].

Here we discuss one of these subnetworks, notably the tetrahydrofolate (THF) system (one-carbon unit pool, C1 pool). This system gives rise to five elementary modes, which are visualized in Fig. 34.1 and are amenable to the following biochemical interpretation. In the mode on the upper left serine is synthesized from glycine by consuming ATP. This is the usual way of serine synthesis in many bacteria, while in higher organisms, it is the other way round in that glycine is produced from serine. The two modes in the middle transfer the C1 unit to dUMP, thus producing dTMP. The C1 group comes either from formate (accompanied by ATP consumption) or from serine. The mode on the lower left corresponds to the formation of formyl-methionine. This is an amino acid specific to bacteria. As it is not present in eukaryotes, *M. pneumoniae* cannot import it from the host. It is also worth noting that the formyl group is stripped off after protein synthesis, which is a nice example of post-translational modification. This cyclic mode is driven by ATP hydrolysis. In another mode realizing formylation of methionine (lower right), the C1 moiety comes from serine.

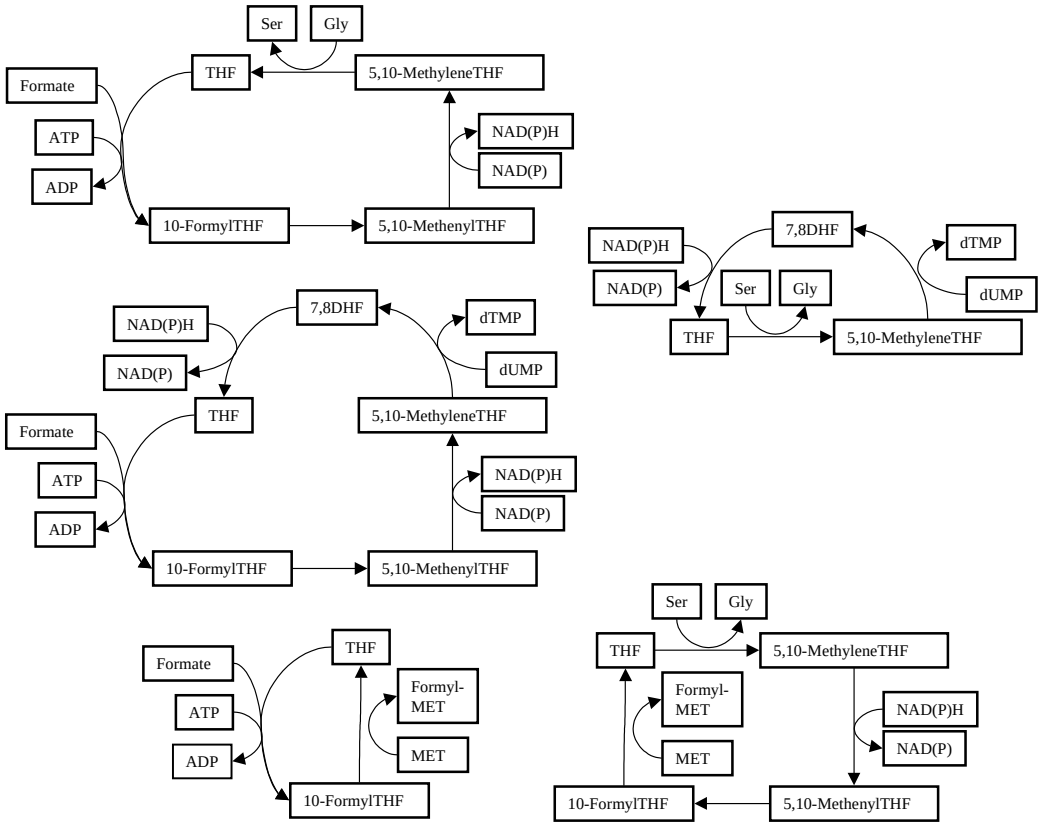

Fig. 34.1 Graphical representation of the five elementary modes in the subnetwork corresponding to THF metabolism in *M. pneumoniae*. Formate, methionyl-tRNA (MET), N-formylmethionyl-tRNA (formyl-MET), serine (Ser), and glycine (Gly) are taken as external metabolites. All cofactors have been made external as well because they participate in more than five reactions in the whole network. All other substances are taken as internal.

As for the THF metabolism in *M. pneumoniae*, many enzyme genes have been functionally assigned by sequence comparison. The complex synthesis of folic acid is not present in this micro-organism and, accordingly, it is resistant against sulfonamides. Note that all the enzymes of this subnetwork are used by at least one elementary mode, except for the reductive steps in which THF is formed from folic acid. These steps do not carry a steady state flux because we have not included the import and export of folic acid. All functionalities of the THF metabolism in *M. pneumoniae* are visible and can be derived directly from the enzyme list without any further intervention, knowledge or data by calculation of the elementary modes of this subnet. Several other subnetworks are considerably less complete in that many enzymes do not enter any mode. This is indicative of missing links in the functional assignment of ORFs. In some instances, pathway analysis shows that the gaps caused by missing enzymes can be bypassed. For ex-

ample, in *M. pneumoniae*, transaldolase seems to be absent. By elementary mode analysis it can be shown that the production of pentoses is still feasible. When such bypasses cannot be detected, the prediction of pathways requires that the hitherto unassigned ORFs are screened for the specific enzyme function that can fill the gap.

Acknowledgement

Financial support by the DFG (Germany) and stimulating discussions with Dr. Juan Carlos Nuño (Madrid) are gratefully acknowledged.

References

1. Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. and Yuan, Y. (1998) Predicting function: From genes to genomes and back, *J. Mol. Biol.* **283**, 707–725.

2. Selkov, E. Jr., Grechkin, Y., Mikhailova, N. and Selkov, E. (1998) MPW: the metabolic pathways database, *Nucl. Acids Res.* **26**, 43–45.

3. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucl. Acids Res.* **27**, 29–34.

4. Dandekar, T., Schuster, S., Snel, B., Huynen, M. and Bork, P. (1999) Pathway alignment: Application to the comparative analysis of glycolytic enzymes, *Biochem. J.* **343**, 115–124.

5. Schilling, C.H., Schuster, S., Palsson, B.O. and Heinrich, R. (1999) Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era, *Biotechnol. Prog.* **15**, 296–303.

6. Schuster S. and Hilgetag C. (1994) On elementary flux modes in biochemical reaction systems at steady state, *J. Biol. Syst.* **2**, 165–182.

7. Schuster, S., Dandekar, T. and Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering, *Trends Biotechnol.* **17**, No. 2, 53–60.

8. Schuster, S., Hilgetag, C. and Fell, D.A. (1994) Detecting elementary modes of functioning in metabolic networks. In *What is Controlling Life?* (Gnaiger, E., Gellerich, F.N. and Wyss, M., eds), pp. 103–105. Innsbruck University Press, Innsbruck.

9. Schuster, S., Fell, D.A., Pfeiffer, T., Dandekar, T. and Bork, P. (1998) Elementary modes analysis illustrated with human red cell metabolism. In *BioThermoKinetics in the Post Genomic Era* (Larsson, C., Påhlman, I.-L. and Gustafsson, L., eds), pp. 332–339. Chalmers, Göteborg.

10. Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C. and Herrmann, R. (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*, *Nucl. Acids Res.* **24**, 4420–4449.

11. Jacobs, E. (1997) Mycoplasma infections of the human respiratory tract. *Wien. Klin. Wochenschr.* **109**, 574–577.

12. Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F. and Schuster, S. (1999) METATOOL: For studying metabolic networks, *Bioinformatics* **15**, 251–257.

13. Mendes, P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3, *Trends Biochem. Sci.* **22**, 361–363.

14. Pfeiffer, T. (1999) Diploma Thesis. Humboldt University Berlin.